DATA ANALYSIS PORTFOLIO

Prepared by

CHUKWUDI JEFFREY

TABLE OF CONTENTS

- Introduction (including project background)
- Root cause analysis process.
- Insights from the dataset,
- Visualisations of these insights
- Important findings and recommendations to solve the business problem.
- Conclusion.

INTRODUCTION

My name is Jeff , I have 6 months experience working as an HR and Data Entry intern at a Logistics Company.

I studied Industrial Mathematics so I enjoy working with numbers, which is why I opted for this course with the Intentions improving my skill in Data Analysis.

INTRODUCTION CONT'D

I was given a few Tables of Data to combine and merge appropriately.

My Job is to Analyse this Data, come up with insights and profer solutions so it can be used to make Data-driven decisions.

ROOT CAUSE ANALYSIS

Q: Why is the Student with the highest score a male ?

A: Because in my experience, higher ratio of males tend to do better than females Academically.

Q: Why do they do better Academically ?

A: Because the Males have more motivation and responsibility.

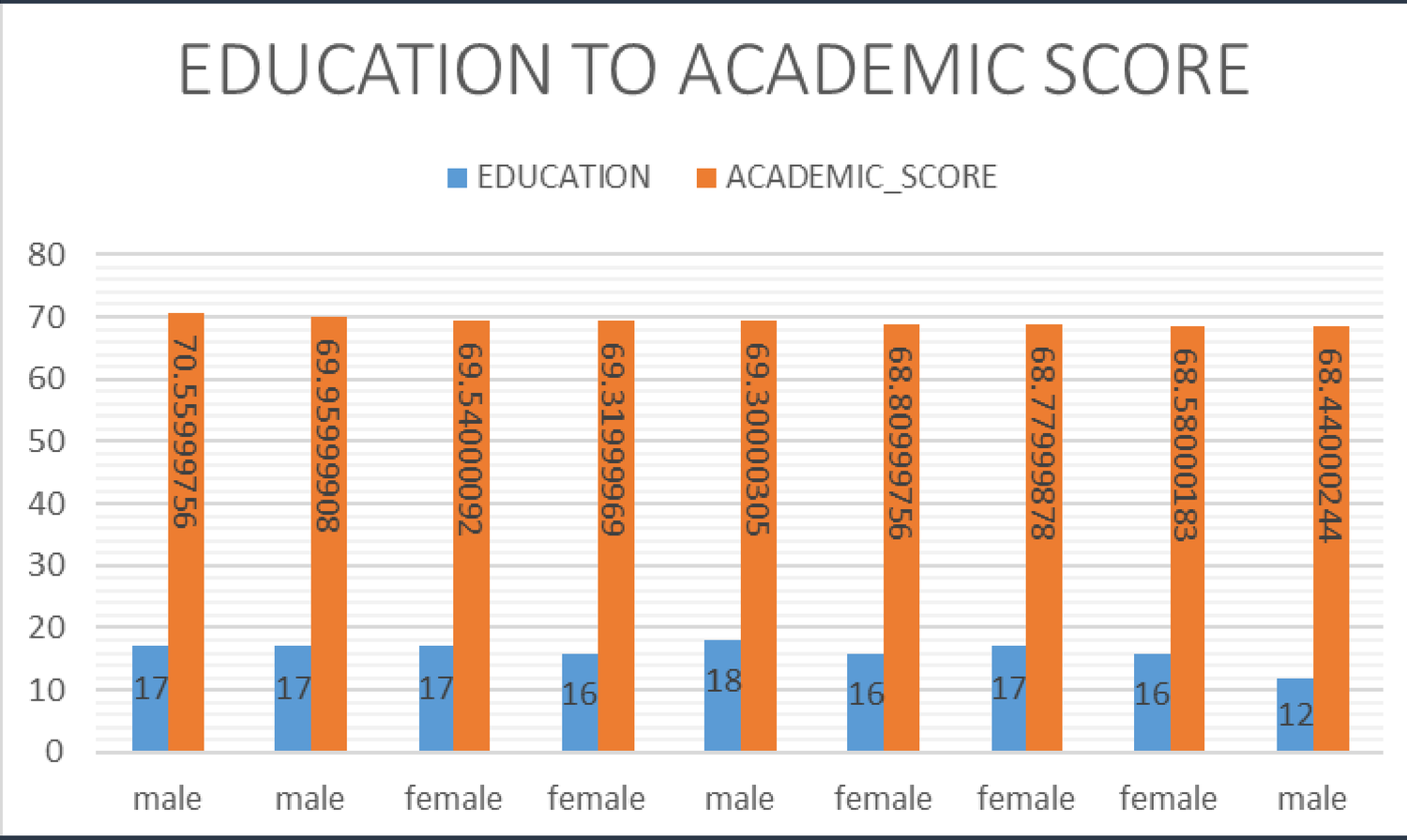
Q: Why do they more responsibilty ?

A: Because they are the providers in their homes.

EDUCATION TO ACADEMIC SCORE

```
SELECT   gender
       ,  education
       ,  academic_score
  FROM   STUDENT_COLLEGE_REPORT
GROUP BY  education, academic_score
ORDER BY  academic_score DESC
```

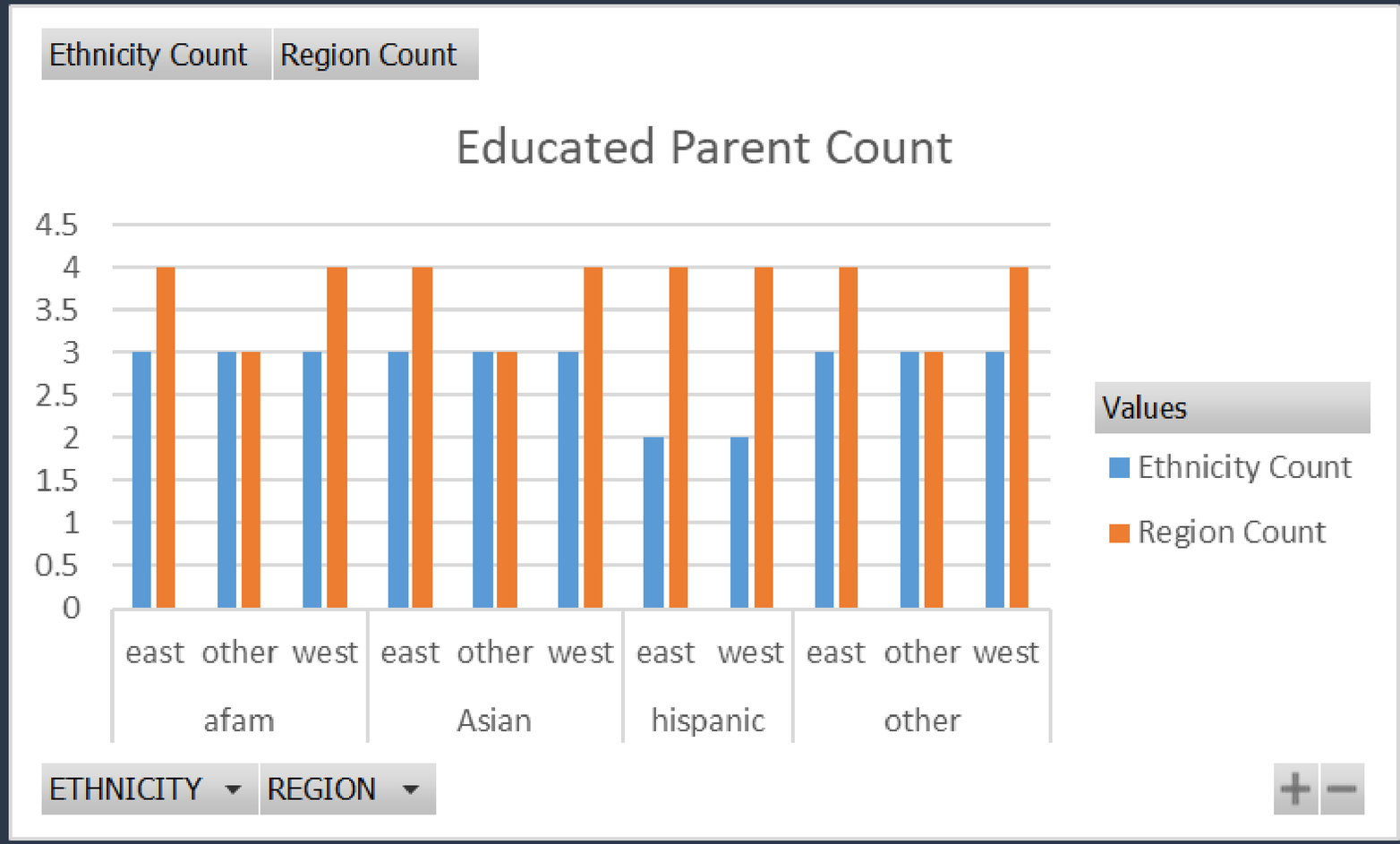
GENDER	EDUCATION	ACADEMIC_SCORE
male	17	70.55999756
male	17	69.95999908
female	17	69.54000092
female	16	69.31999969
male	18	69.30000305
female	16	68.80999756
female	17	68.77999878
female	16	68.58000183
male	12	68.44000244



NUMBER OF EDUCATED PARENTS

```
SELECT    ethnicity
        ,   count(ethnicity) over (partition by ethnicity) ETHCOUNT
        ,   region
        ,   count() over (partition by region) REGCOUNT
  FROM   Student_college_report
 WHERE   fcollege = 'yes' AND mcollege = 'yes'
GROUP BY  ethnicity, region  ;
```

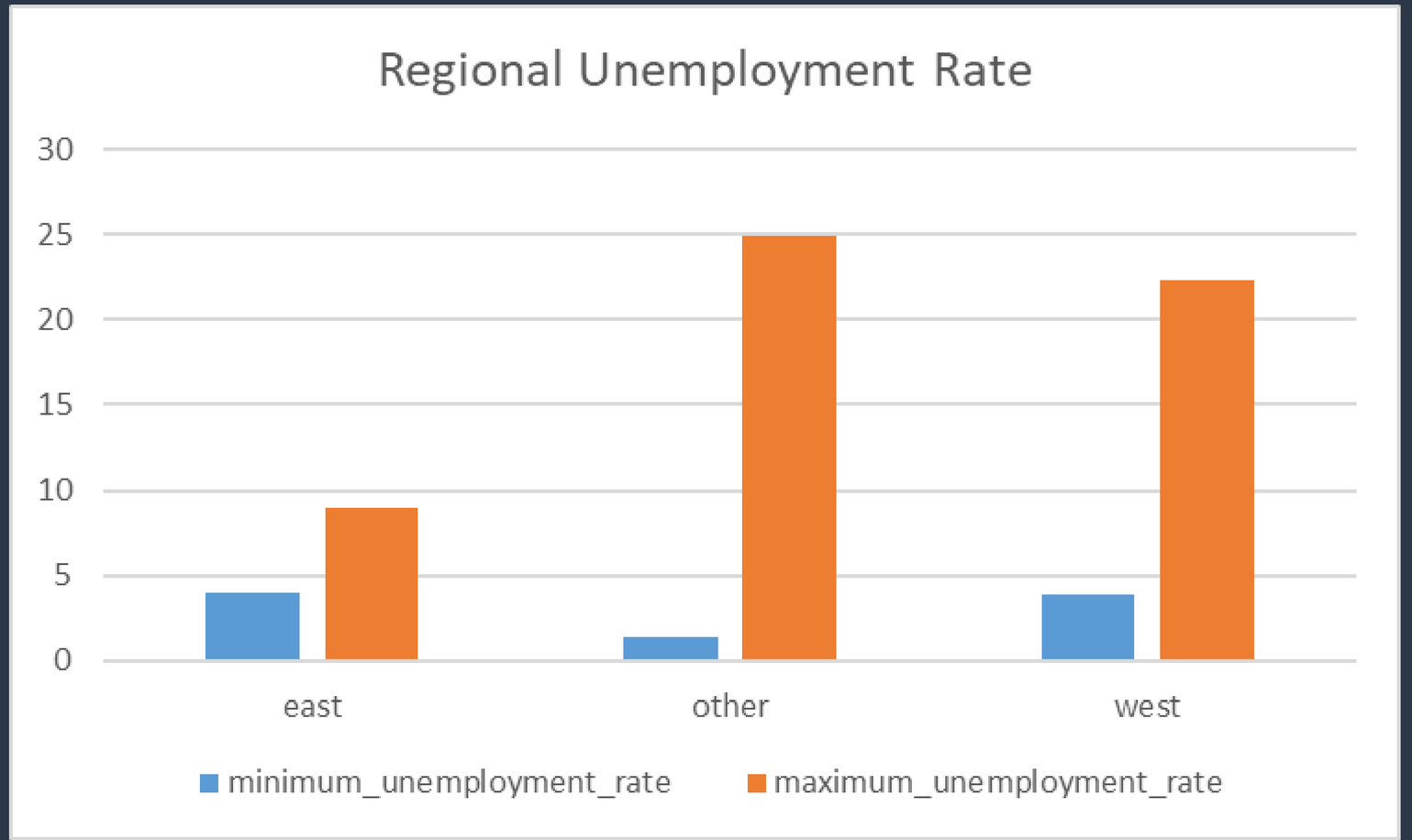
ETHNICITY	REGION	Ethnicity Count	Region Count
afam	east	3	4
afam	other	3	3
afam	west	3	4
Asian	east	3	4
Asian	other	3	3
Asian	west	3	4
hispanic	east	2	4
hispanic	west	2	4
other	east	3	4
other	other	3	3
other	west	3	4



REGIONAL UNEMPLOYMENT RATE

```
SELECT   region
       ,   min(unemp) minimum_unemployment_rate
       ,   max(unemp) maximum_unemployment_rate
 FROM   (    SELECT   region
                    ,   unemp
               FROM   Student_college_report
            GROUP BY   region
                    ,   unemp)
GROUP BY  region ;
```

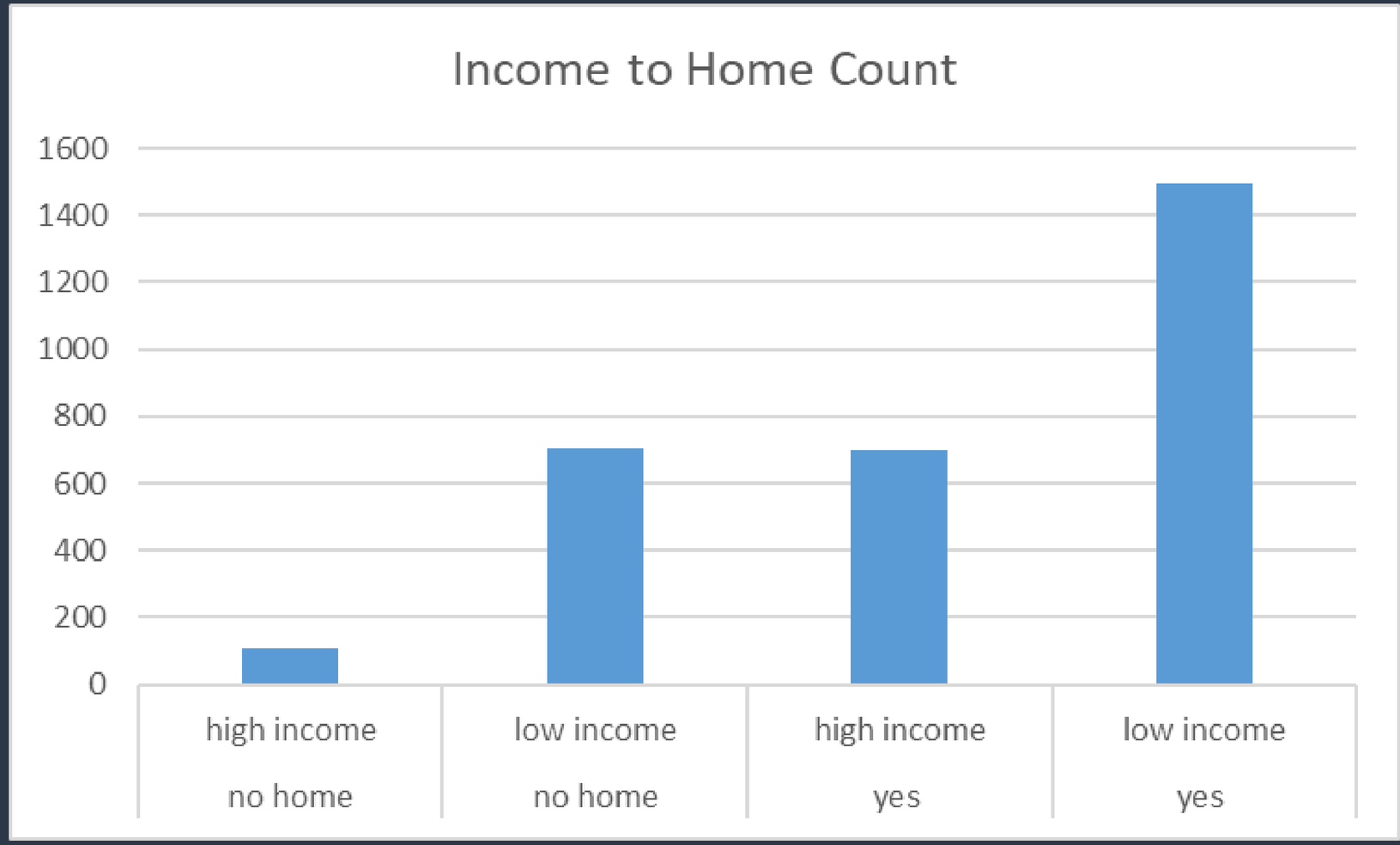
REGION	MINIMUM_UNEMPLOYMENT_RATE	MAXIMUM_UNEMPLOYMENT_RATE
east	4	9
other	1.399999976	24.89999962
west	3.900000095	22.29999924



HOME AND INCOME COUNT

```
SELECT   home
       ,  income
       ,  COUNT()
  FROM   STUDENT_COLLEGE_REPORT
GROUP BY  home, income  ;
```

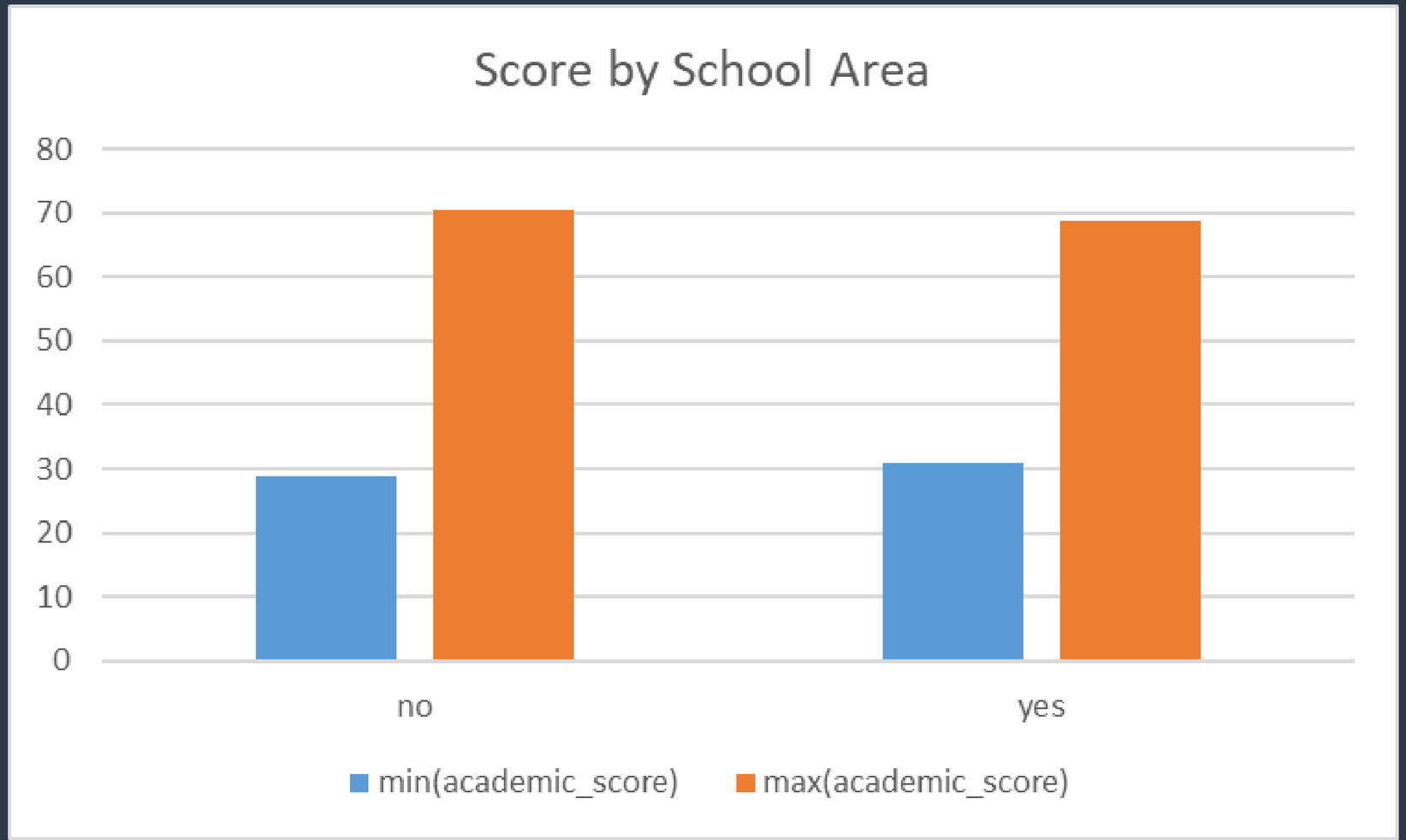
HOME	INCOME	COUNT
no home	high income	110
no home	low income	707
yes	high income	700
yes	low income	1495



URBAN TO ACADEMIC SCORE

```
SELECT   urban
       ,  min(academic_score)
       ,  max(academic_score)
  FROM  (SELECT     URBAN
                  ,    academic_score
                FROM    STUDENT_COLLEGE_REPORT
              GROUP BY  URBAN, ACADEMIC_SCORE
              ORDER BY   ACADEMIC_SCORE DESC)
GROUP BY  urban ;
```

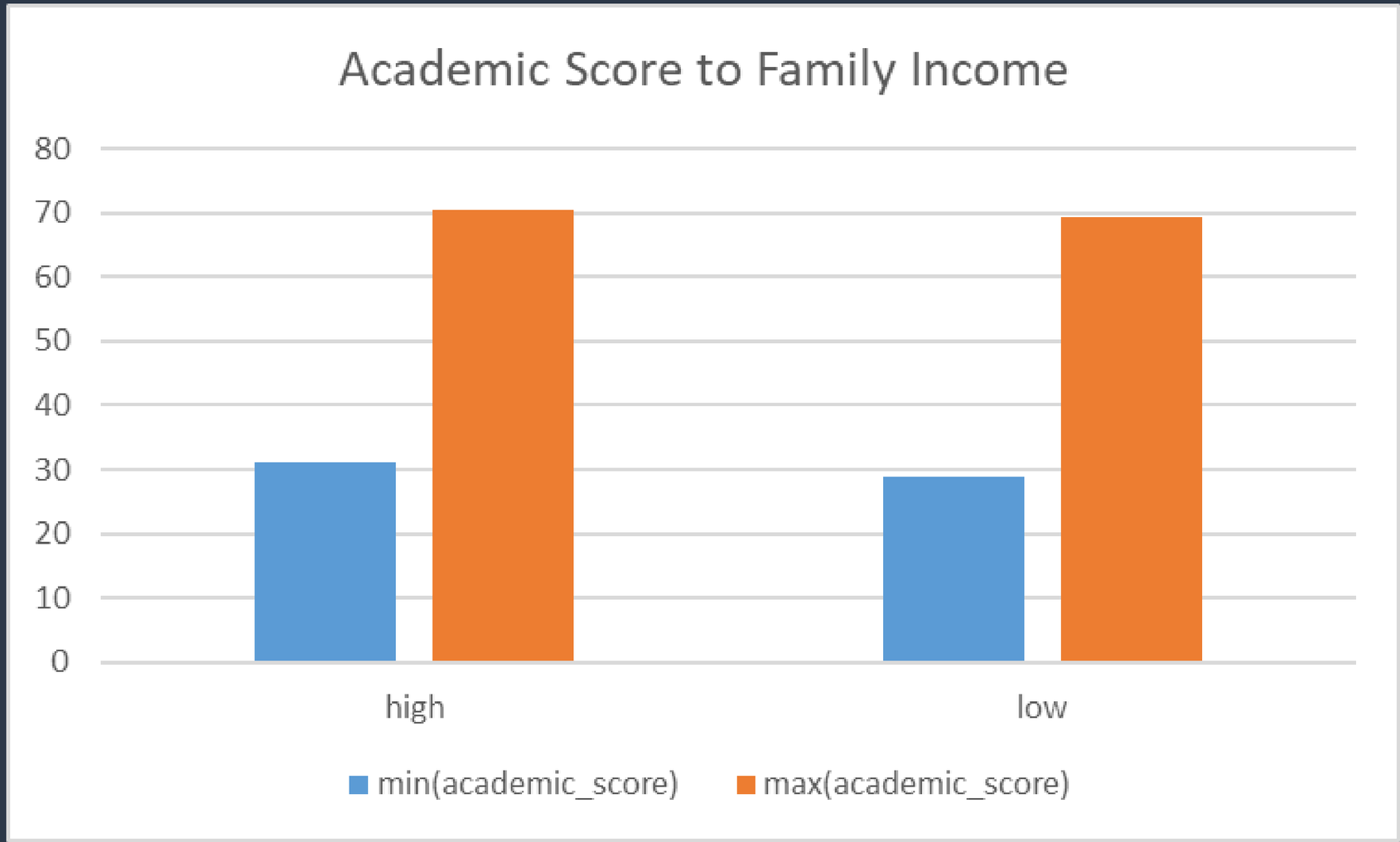
URBAN	MIN(ACADEMIC_SCORE)	MAX(ACADEMIC_SCORE)
no	28.95000076	70.55999756
yes	30.97999954	68.80999756



INCOME TO ACADEMIC SCORE

```
SELECT   income
       ,   min(academic_score)
       ,   max(academic_score)
  FROM   (SELECT   income
                 ,   academic_score
            FROM   Student_college_report
        GROUP BY  income, academic_score
        ORDER BY  academic_score DESC)
GROUP BY income ;
```

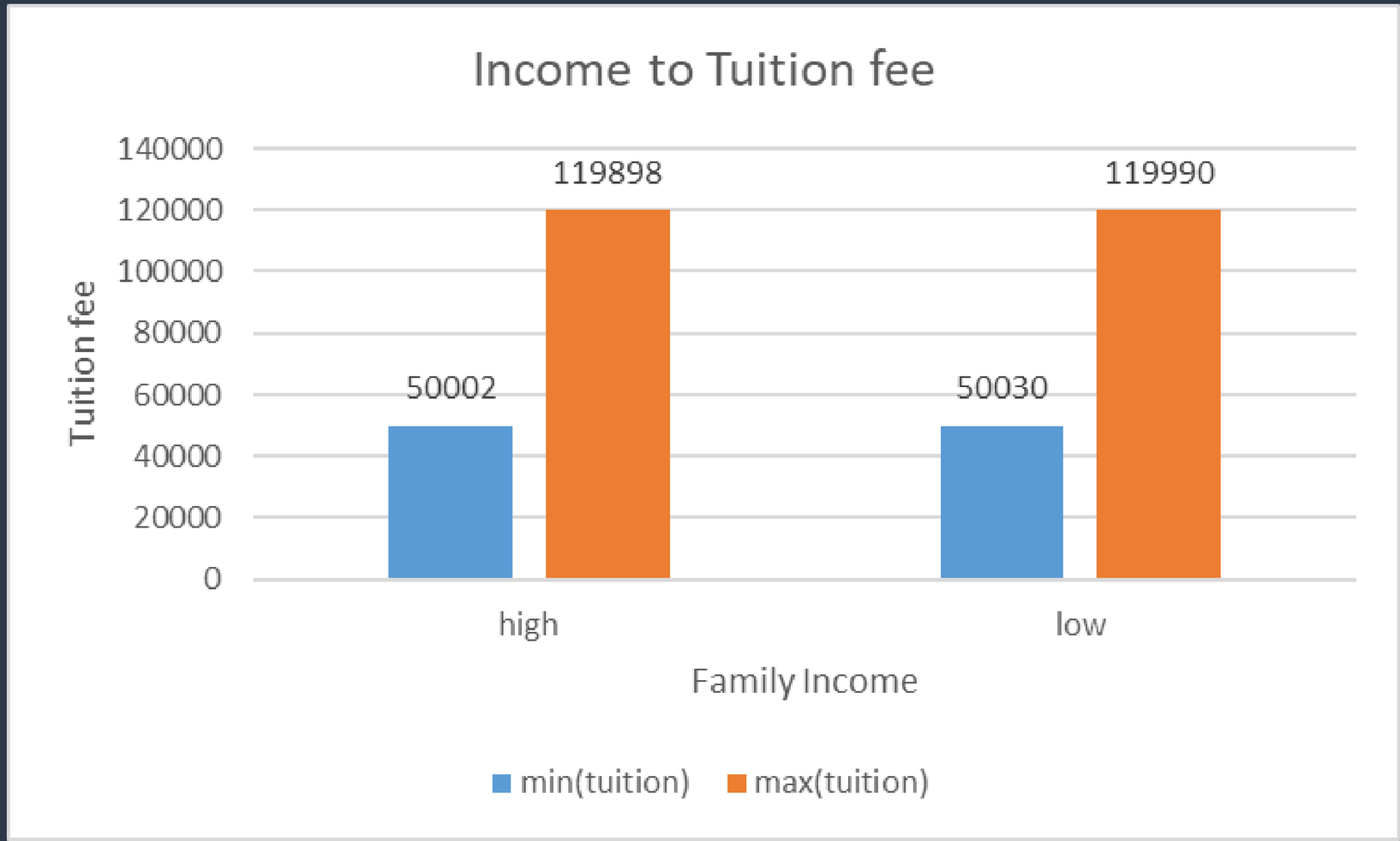
INCOME	MIN(ACADEMIC_SCORE)	MAX(ACADEMIC_SCORE)
high	31.04999924	70.55999756
low	28.95000076	69.31999969



INCOME TO TUITION FEE

```
SELECT   income
         ,   min(tuition)
         ,   max(tuition)
 FROM    (   SELECT  income
                     ,  tuition
               FROM   Student_college_report
             GROUP BY   income, tuition
             ORDER BY   tuition DESC)
GROUP BY  income ;
;
```

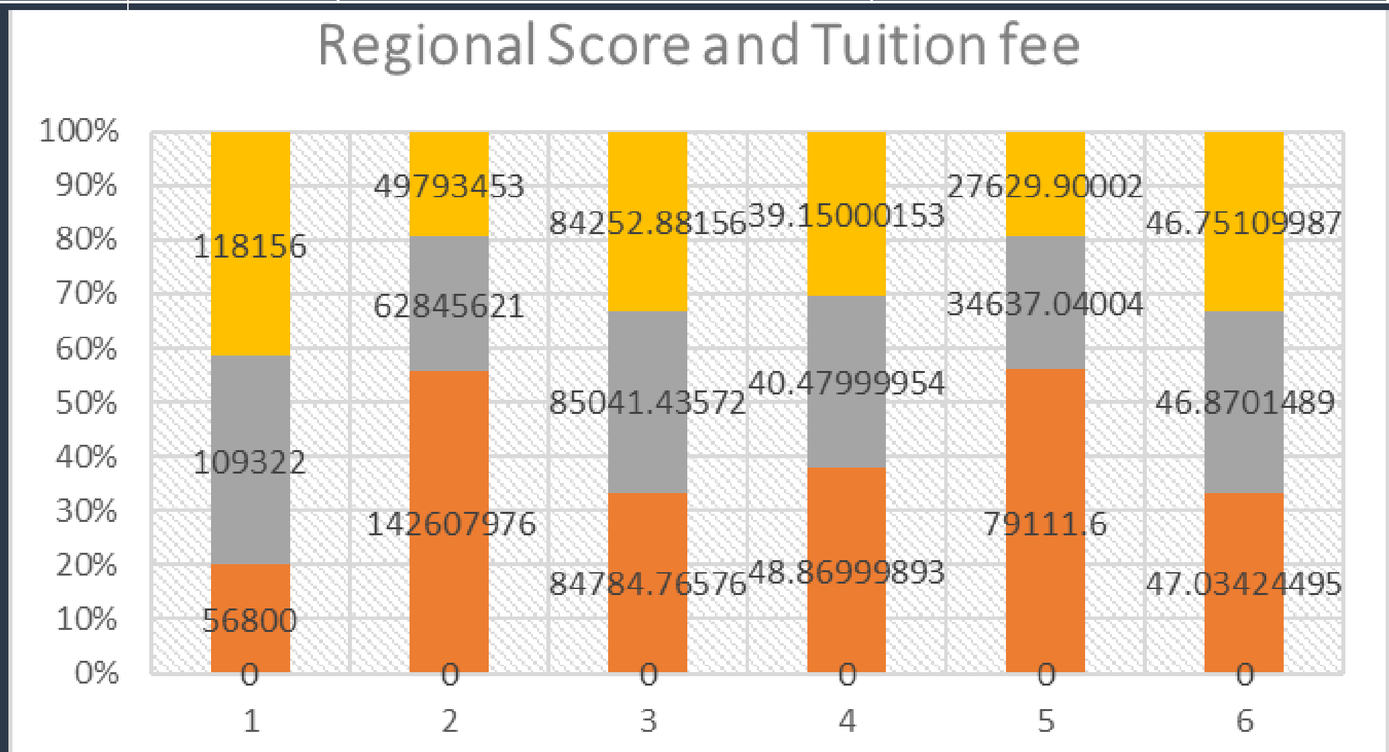
Income	min(tuition)	max(tuition)
high	50002	119898
low	50030	119990



REGIONAL SCORE AND TUITION FEE

```
SELECT   region
       ,  tuition
       ,  sum(tuition)
       ,  avg(tuition)
       ,  academic_score
       ,  sum(academic_score)
       ,  avg(academic_score)
  FROM   Student_college_report
GROUP BY  region
ORDER BY  tuition, academic_score ;
```

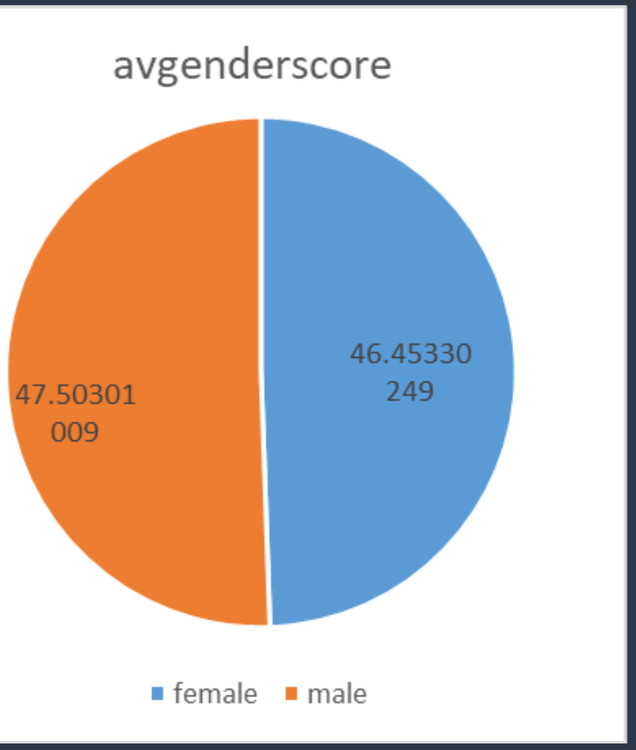
REGION	TUITION	TUITION SUM	TUITION AVERAGE	ACADEMIC SCORE	SUM OF ACADEMIC SCORE	AVERAGE ACADEMIC SCORE
other	56800	142607976	84784.76576	48.86999893	79111.6	47.03424495
east	109322	62845621	85041.43572	40.47999954	34637.04004	46.8701489
west	118156	49793453	84252.88156	39.15000153	27629.90002	46.75109987

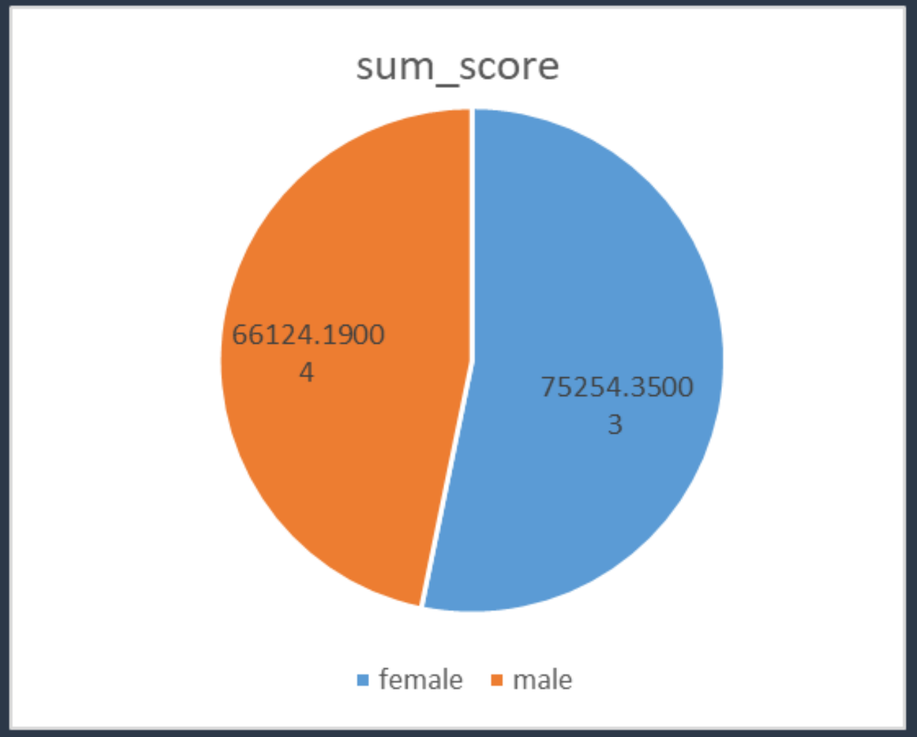


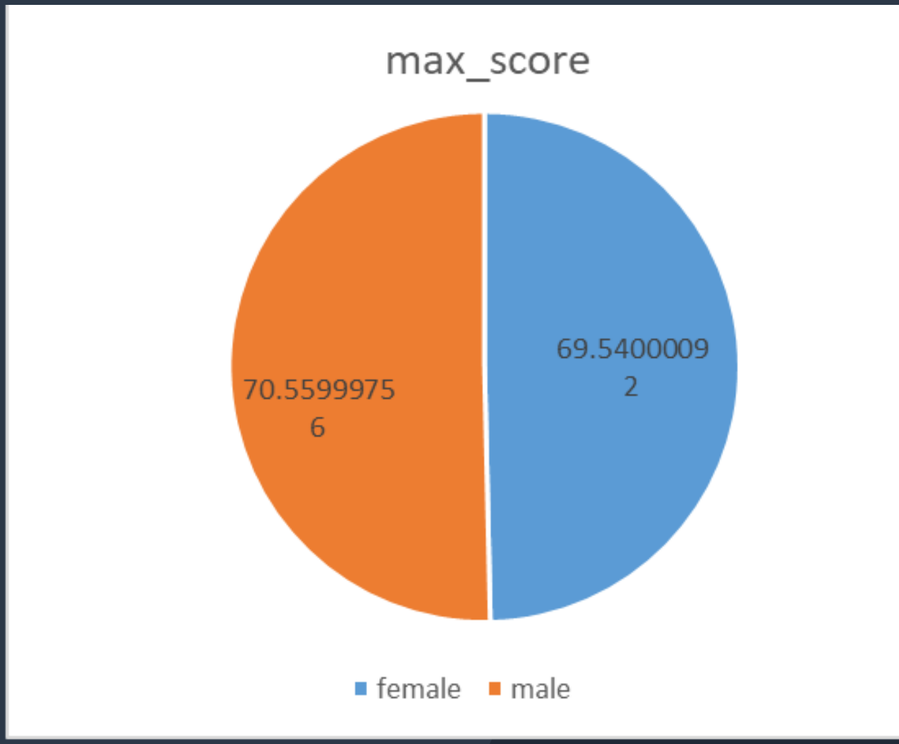
GENDER AVERAGE, SUM AND MAXIMUM SCORE

```
SELECT   gender
       ,  avg(academic_score)
       ,  sum(academic_score)
       ,  max(academic_score)
  FROM   STUDENT_COLLEGE_REPORT
GROUP BY  GENDER ;
```

GENDER	AVERAGE GENDER SCORE	SUM OF SCORE	MAXIMUM SCORE
FEMALE	46.45330249	75254.35003	69.54000092
MALE	47.50301009	66124.19004	70.55999756



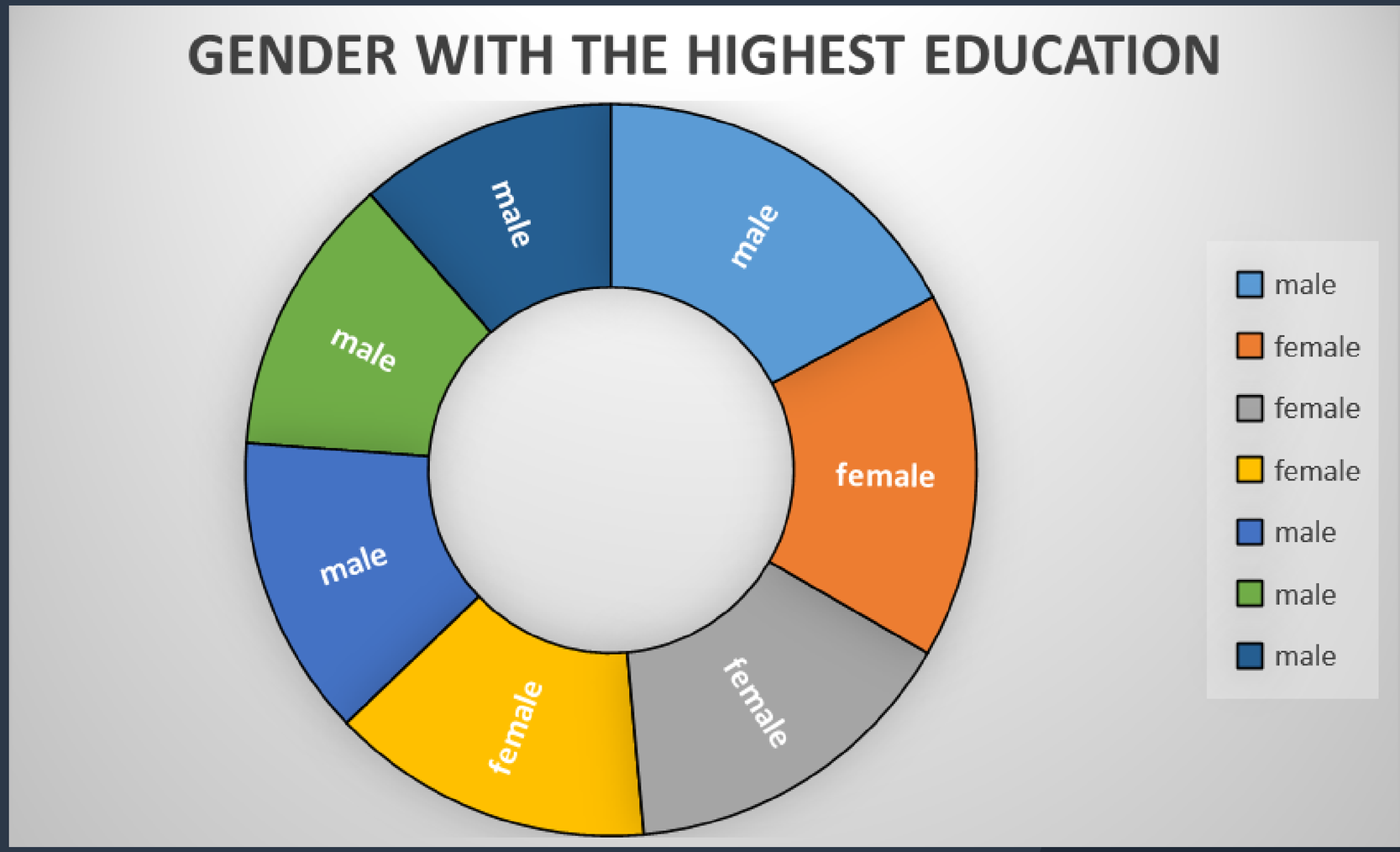




GENDER WITH THE HIGHEST EDUCATION

```
SELECT  gender
        ,  education
  FROM  Student_college_report
GROUP BY  education
ORDER BY   education DESC ;
```

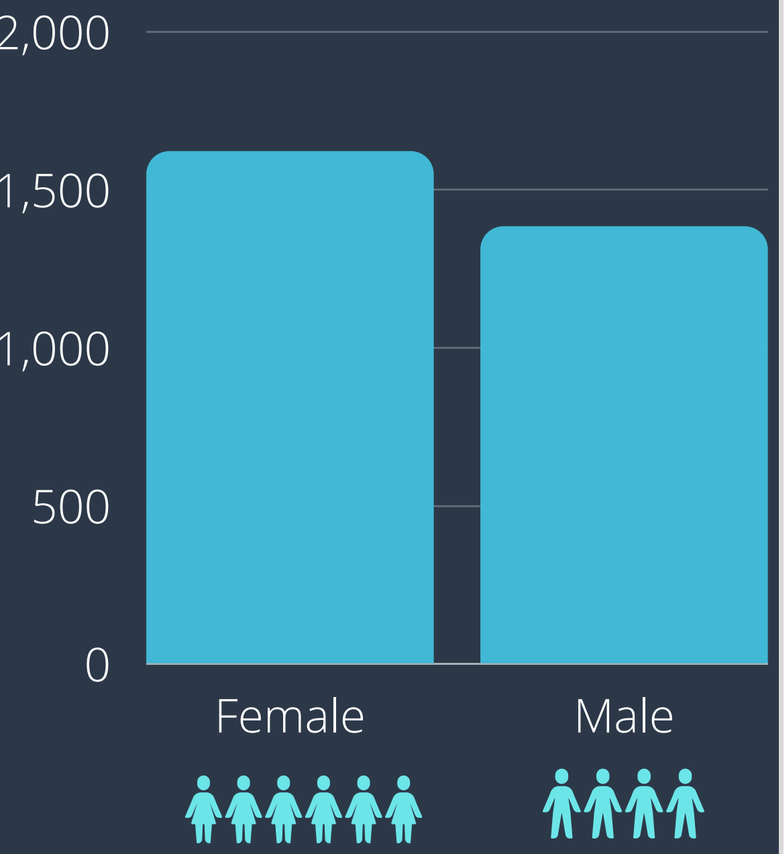
GENDER	COUNT
male	18
female	17
female	16
female	15
male	14
male	13
male	12

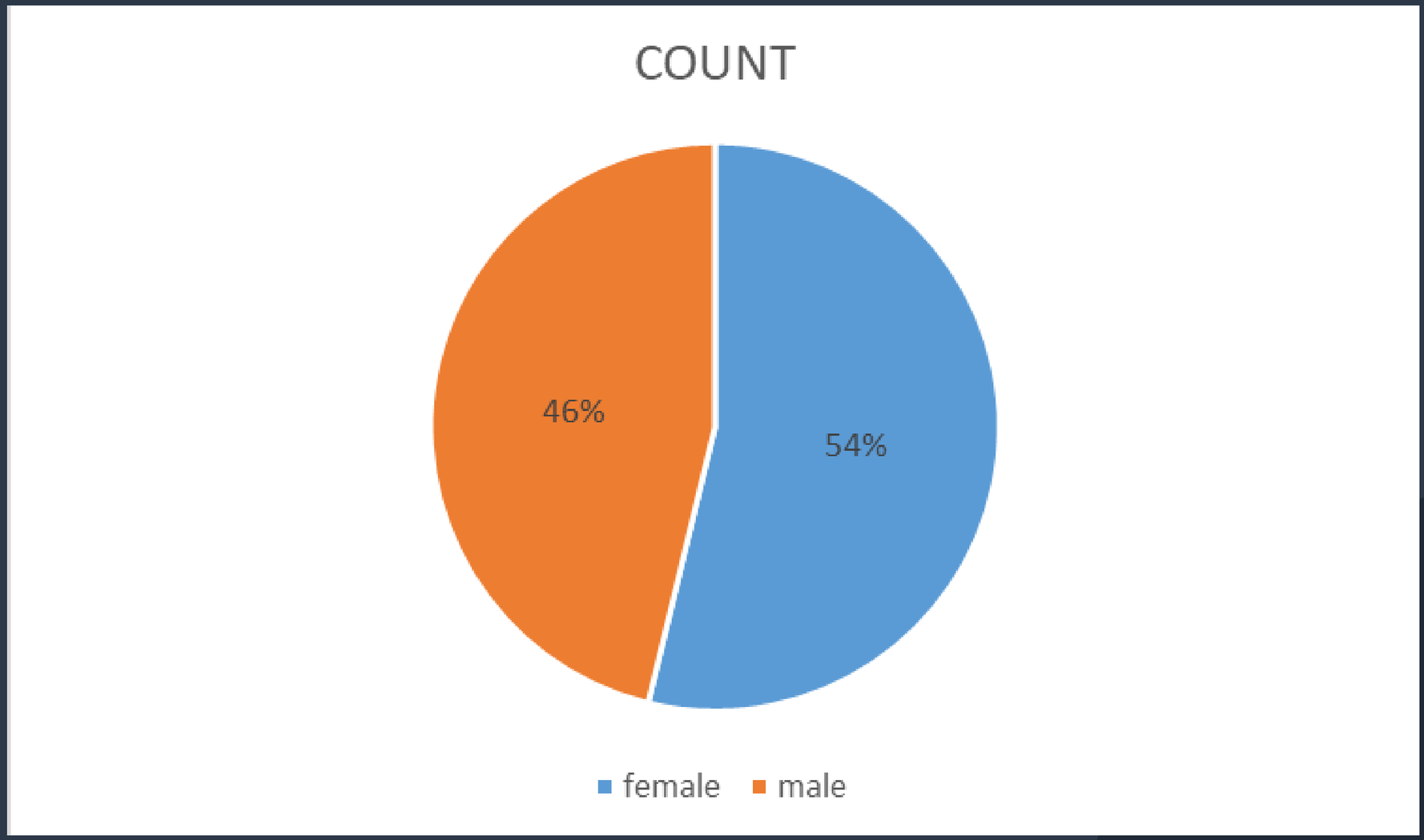


GENDER COUNT

```
SELECT  gender
      ,  COUNT()
  FROM  STUDENT_COLLEGE_REPORT
GROUP BY  GENDER
  HAVING  COUNT()
ORDER BY  COUNT() DESC
                    ;
```

GENDER	COUNT
female	1620
male	1392

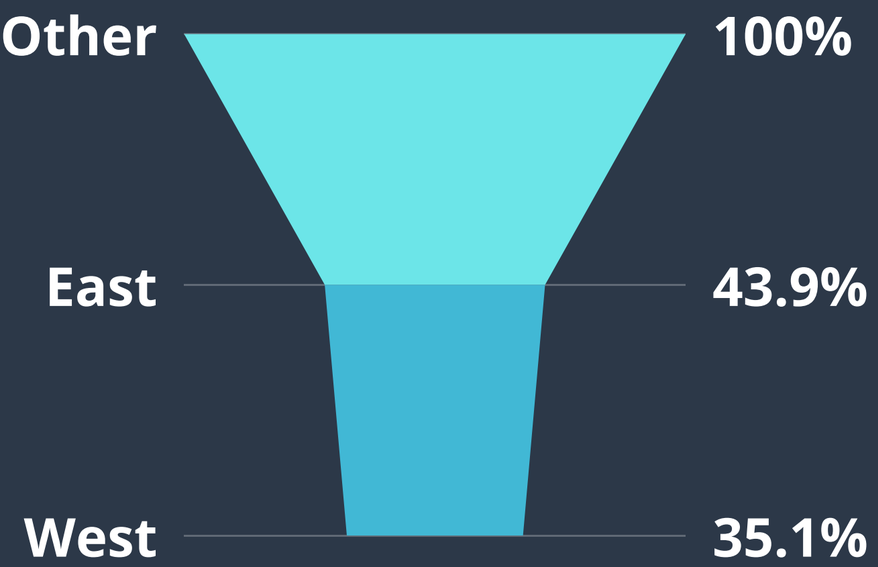


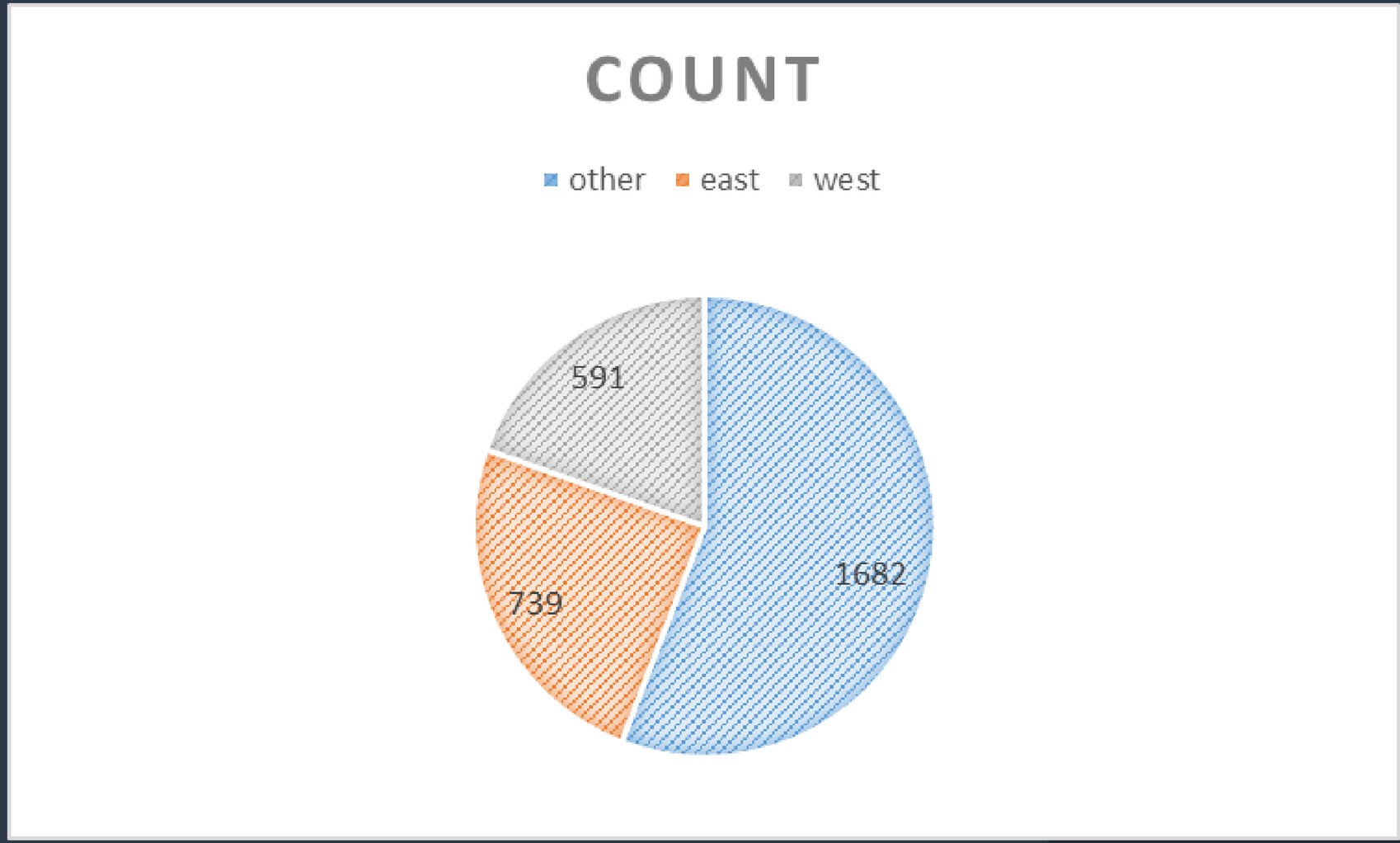


REGIONAL COUNT

```
SELECT   region
       ,  COUNT()
  FROM   STUDENT_COLLEGE_REPORT
GROUP BY   REGION
  HAVING   COUNT()
ORDER BY   COUNT() DESC
                    ;
```

REGION	COUNT
other	1682
east	739
west	591

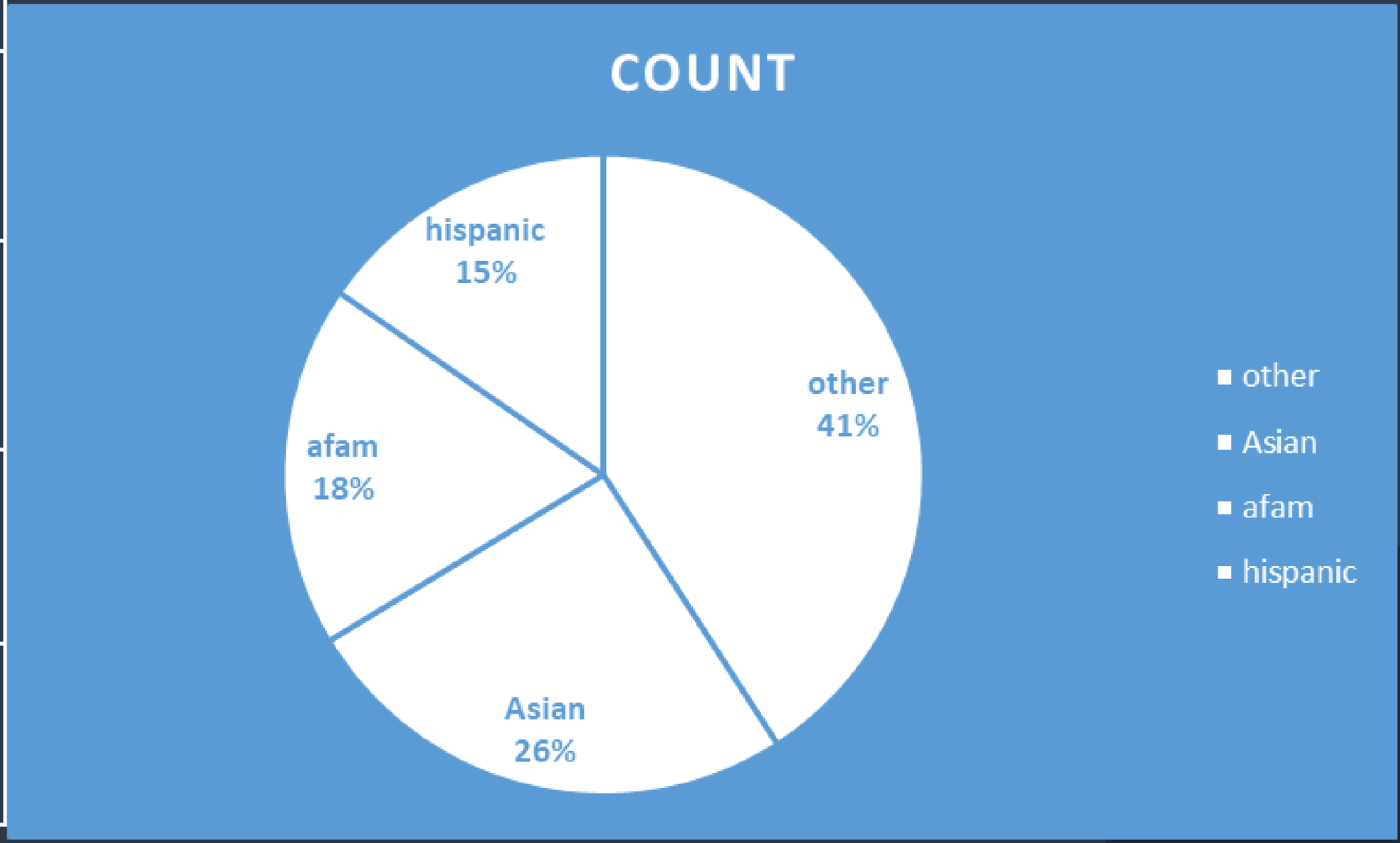




ETHNICITY COUNT

```
SELECT   ethnicity
       , COUNT(ethnicity)
  FROM   STUDENT_COLLEGE_REPORT
GROUP BY ETHNICITY
HAVING   COUNT()
ORDER BY COUNT() DESC
                              ;
```

ETHNICITY	COUNT
Other	1231
Asian	767
Afam	551
Hispanic	463



FINDINGS AND RECOMMENDATIONS

- From the above Analysis, we discover that the female population is higher than the male population.
- We also discover that the highest "Academic score" isn't the person with the most Education years and it's a male.
- The ethnicity with the most population in the Dataset is "Asian" and the East has more population than the west.
- We also noticed that the Asian population has more Educated parents than the rest.
- Families with 'High' income paid lesser tuition fee for their students than the families with 'Low' income.

FINDINGS AND RECOMMENDATIONS CONT'D

- We also notice that the student with the highest 'Academic score' is from a 'high income' home.
- The rate of 'Unemployment' is higher in the 'West'.
- We also notice that families with 'low income' that have their own homes are higher in population than the families with 'high income'.
- We notice that Schools located in urban areas have a lower 'Maximum Academic score' than Schools not located in Urban areas.

CONCLUSION

The insights drawn from the Dataset, the most populated ethnicity is "ASIAN" and the most populous region is the "EAST".

Also, the rate of unemployment is higher in the "WEST" and families that have their own homes with "LOW INCOME" are more then families without their own home and have a "HIGH INCOME".

Finally, we discovered that the person with the highest score is a "MALE" and he only has 17 years Education experience.

THANK YOU!

E-mail : jeffboytwist@gmail.com
Linked In: http://www.linkedin.com/in/jeffrey-twist-a06593202